



## Fields From Markov Chains

**Justesen, Jørn**

*Published in:*  
I E E E Transactions on Information Theory

*Link to article, DOI:*  
[10.1109/TIT.2005.858988](https://doi.org/10.1109/TIT.2005.858988)

*Publication date:*  
2005

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Justesen, J. (2005). Fields From Markov Chains. *I E E E Transactions on Information Theory*, 51(12), 4358-4362. <https://doi.org/10.1109/TIT.2005.858988>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

where  $\omega$  is now changed to a complex, primitive  $p^{(K+1)}$ th root of unity. Two examples on rotated AM-PSK constellations are shown in Fig. 2. With some straightforward adaptations of the proof of Theorem 2, it can be shown that the mapper  $\mu$  defined by (8) is rank-distance preserving as well.

#### APPENDIX

Lemmas that are used in proving Theorem 2 are provided here for reference.

**Lemma 3 ([2]):** Let  $H$  be an  $(m \times n)$  matrix with components drawn from a subfield  $\mathbb{F}$  of the complex numbers  $\mathbb{C}$ . If  $\underline{a}^t H = \underline{0}^t$ ,  $\underline{a} \in \mathbb{F}^m$ , implies  $\underline{a} = \underline{0}$ , then  $H$  has rank  $m$  even when considered as a matrix over the complex numbers.  $\square$

**Lemma 4 ([2]):** Let  $p$  be a prime and let  $\omega$  be a complex, primitive  $p^K$ th root of unity  $K \geq 1$ . Then

$$\mathbb{Z}[\omega]/(1 - \omega) \cong \mathbb{F}_p. \quad \square$$

**Lemma 5 ([2]):** Let  $p$  be a prime and let  $\omega$  be a complex, primitive  $p^K$ th root of unity; then

$$\Delta s = \frac{\omega \sum_{k=0}^{K-1} p^k a_k - \omega \sum_{k=0}^{K-1} p^k b_k}{\omega - 1} \in \mathbb{Z}[\omega] \quad (9)$$

where  $a_k, b_k \in \mathbb{Z}_p$ . Moreover

$$\Delta s \equiv a_0 \ominus_p b_0 \pmod{1 - \omega} \quad (10)$$

where  $\ominus_p$  denotes subtraction in  $\mathbb{F}_p$ .  $\square$

#### REFERENCES

- [1] E. Gabidulin, "Theory of codes with maximum rank distance," *Probl. Inf. Transm.*, vol. 21, pp. 3–16, Jan./Mar. 1985.
- [2] H. F. Lu and P. V. Kumar, "A unified construction of space-time codes with optimal rate-diversity tradeoff," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1709–1730, May 2005.
- [3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
- [4] H. F. Lu and P. V. Kumar, "Rate-diversity tradeoff of space-time codes with fixed alphabet and optimal constructions for PSK modulation," *IEEE Trans. Inf. Theory*, no. 10, pp. 2747–2751, Oct. 2003.
- [5] B. J. Belzer, A. D. Risley, P. Hou, and T. R. Fischer, "Capacity of AM-PSK on partially coherent fading channels," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 369–373, Mar. 2002.
- [6] X. Dong, N. C. Beaulieu, and P. H. Wittke, "Signaling constellations for fading channels," *IEEE Trans. Commun.*, vol. 47, no. 5, pp. 703–714, May 1999.
- [7] A. R. Hammons Jr., "New AM-PSK space-time codes achieving the rate-diversity tradeoff," in *Proc. 2004 Int. Symp. Information Theory and Its Applications*, Parma, Italy, Oct. 2004, pp. 320–324.
- [8] —, "Space-time codes that achieve the rate-diversity tradeoff for certain AM-PSK modulations," *IEEE Trans. Inf. Theory*, submitted for publication.

## Fields From Markov Chains

Jørn Justesen, *Member, IEEE*

**Abstract**—A simple construction of two-dimensional (2-D) fields is presented. Rows and columns are outcomes of the same Markov chain. The entropy can be calculated explicitly.

**Index Terms**—Markov chain, Markov random field, two-dimensional (2-D) field.

#### I. INTRODUCTION

There are extremely few two-dimensional (2-D) discrete fields that permit a simple analysis. In particular it is rarely possible to calculate the entropy explicitly or to find probabilities of finite symbol configurations. We present a simple construction that allows rows and columns to be outcomes of the same Markov chain. Such fields may be useful in constructing codes for two-dimensional storage, or they may serve as models of images. One may think of these fields as crosswords in a regular language [1], [2].

The most important class of 2-D fields that admits a detailed analysis is the Pickard fields [3], [4]. A discussion of Pickard fields and other Markov random fields on finite rectangular lattices was given in [5]. However, important questions about the application of these results remain unanswered, and in particular the structure of fields with constraints on the symbol sequences was not considered in [5].

Constrained arrays are often considered as models of 2-D information sources and as adaptations to the properties of recording media [1], [6]. In Section II we give some definitions relating to constrained arrays and discuss the relation to crosswords. In the rest of the paper it is assumed that the constraints can be satisfied by requiring the rows and columns of the array to follow the probability distribution of a Markov chain. If a finite alphabet Markov chain is given, the construction described in Section III can be applied directly.

#### II. CONSTRAINED TWO-DIMENSIONAL ARRAYS

We consider the set of  $N$  by  $L$  arrays,  $\Psi(N, L)$ , where the entries,  $b(r, s)$ , are taken from a finite alphabet  $A = \{a_0, a_1, \dots, a_J\}$ . We require that the arrays satisfy the following constraint: In any row or column, any two consecutive symbols must belong to a set of admissible pairs,  $\Omega$ . If the number of such arrays is  $|\Psi(N, L)|$ , the *combinatorial entropy* of the constraint is defined as

$$C = \lim_{N, L \rightarrow \infty} \left\{ \frac{\log |\Psi(N, L)|}{NL} \right\}. \quad (1)$$

Thus,  $C$  measures how the number of arrays increases with their size.

Shannon gave a well-known solution to the problem in one dimension [1]. For the given constraint,  $C$  may be found as the logarithm of the maximal eigenvalue of the adjacency matrix indicating the permissible transitions. Moreover, if the sequences are converted to a Markov chain by using a suitable assignment of transition probabilities, the maximal entropy of the Markov chain equals the combinatorial entropy. In two dimensions, the properties of fields are much more complex. When a field consistent with an admissible subset can be constructed,

Manuscript received August 26, 2004; revised April 23, 2005.

The author is with the COM, Technical University of Denmark, DK 2800 Kgs Lyngby, Denmark (e-mail: jju@com.dtu.dk).

Communicated by V. A. Vaishampayan, Associate Editor At Large.

Digital Object Identifier 10.1109/TIT.2005.858988

the maximum entropy is attained by a Markov random field. However, it is in general difficult to find the combinatorial entropy.

In [1] the authors discussed Shannon's statement (published without proof) of a relation between the entropy of a language and the possibility of constructing crosswords. Crosswords may be seen as a particular instance of the above problem, however in the early days of information theory, the concepts of formal languages and grammars were not available. In [2] natural languages were modeled as Markov chains (both on the word and the sentence level), and this approach was the basis for assigning an entropy to languages.

The definition of a formal language does not involve probabilities, although later probabilities may be assigned to the derivations. Similarly other types of constrained finite alphabet systems are often initially defined without a probability distribution. We discuss only the simplest kind of formal languages, finite state or regular languages. In particular we do not discuss languages that are designed to describe two-dimensional structures. Instead we simplify the general regular languages to follow the definition given above: A word may start and end with arbitrary symbols, and the sequence of symbols has to agree with the set of admissible pairs. We assume that any admissible string can be extended to an arbitrarily long word in the language. If we start from a language generating finite length words, we can concatenate these by considering the closure of the language, and if desired we can add a space to separate words. However, we do not require that the limits of the finite rectangle coincide with the beginning or end of such finite words.

*Example 1:* For the alphabet  $\{C, O, D, E\}$ , let the set of admissible pairs be

$$\Omega = \{(CC), (CO), (CE), (OD)(OE), (DE), (EC), (ED)\}.$$

Thus, *DECODE* is a word.

We do not assume that the language is symmetrical, and thus, it must be specified that the description applies to the sequences of symbols read left to right or top to bottom. The indexing of the array is chosen to agree with the indexing of the Markov chain. However, it is important that the model is developed in order of increasing indices in one dimension, but in the opposite order in the other dimension. Thus, the description will be left-to-right but bottom-up. For symmetric languages we can use the common top-down ordering of the symbols. To generate an  $N$  by  $L$  crossword, we first fill in the left column, and next the bottom row starting with the left symbol. The following rows are filled in one symbol at a time starting from the bottom and working left to right.

*Lemma 1:* If all pairs  $(b(r, s-1), b(r, s))$  and  $(b(r, s), b(r+1, s))$  are in  $\Omega$  for  $r > r'$  and for  $r = r'$  and  $s < s'$ , then there is at least one possible choice for  $b(r', s')$ .

*Proof:*  $(b(r', s'-1), b(r', s'))$  and  $(b(r', s'), b(r'+1, s'))$  must be in  $\Omega$ . Since  $b(r'+1, s'-1)$  satisfies the same constraint, we can always choose the same symbol for  $b(r', s')$ .

Lemma 1 gives the reason for specifying the symbols in this slightly unnatural order. If the array is generated left to right starting from the top,  $b(r, s)$  is constrained to follow the symbols  $b(r, s-1)$  and  $b(r-1, s)$ . However, there may not be a symbol that satisfies both of these conditions, and consequently some previous choices may have to be changed.

For a particular choice of the last row, we can get a possible row above by shifting all symbols to the right. In this way all rows and columns become shifts of the same word.

Thus, any word of length  $N + L - 1$  can be converted into a trivial array, and the set of arrays is not empty.

*Corollary 1:* If the trivial arrays are the only solutions, the combinatorial entropy is zero, otherwise it is positive.

*Example 1 (Continued):* If an array were generated left to right and from the top, we might end in a situation as

$$\begin{array}{cc} O & E \\ D & ? \end{array}$$

which cannot be completed. On the other hand

$$\begin{array}{cc} D & ? \\ E & C \end{array}$$

can be completed by  $E$  (and in this case it is the only possibility).

*Lemma 2:* The causal (symbol by symbol) generation of arrays generates all arrays that satisfy the given constraint.

*Proof:* The initial column and row include all admissible possibilities. In the subsequent choices, a symbol is only excluded if it violates either the row or column constraint.

*Example 1 (Continued):* Starting with the left column and the bottom row, we generate a 4 by 4 array:

$$\begin{array}{cccc} C & & & \\ O & & & \\ D & & & \\ E & C & C & E \end{array}$$

From  $\Omega$  we get that  $b(3, 2)$  must be  $E$ . Since  $b(3, 3)$  follows  $E$ , it can be  $C$  or  $D$ , but  $C$  cannot follow  $D$ . The symbol  $b(3, 4)$  follows a  $C$ , and could, thus, be  $C$ ,  $O$ , or  $E$ . But since it is followed by  $E$ , it cannot be an  $E$ . Thus, the third row can be  $DECO$  or  $DECC$ . By similarly filling in row 2 and 1 from left to right we may get

$$\begin{array}{cccc} C & E & D & E \\ O & D & E & C \\ D & E & C & O \\ E & C & C & E \end{array}$$

*Corollary 2:* A nontrivial crossword based on a regular language exists if and only if there is at least one pair of symbols  $(a_u, a_v)$  such that for  $b(r, s-1) = a_u$  and  $b(r, s+1) = a_v$ , there are at least two admissible values for  $b(r, s)$ .

Thus, for languages with this simple structure (which may, however, serve as a first approximation to natural languages), crosswords are possible even in some cases where the entropy of the language is low. Clearly Shannon's assumption of "a rather chaotic and random nature" is not satisfied.

### III. THE PROBABILITY DISTRIBUTION

We assume that a regular, irreducible finite state Markov chain  $M$  is given. Thus, the state at point  $(r, s)$  is a random variable  $B(r, s)$  with value  $b(r, s)$  in the finite alphabet  $A$ . The stationary probability of state  $j$  is  $Q_j = P[b(r, s) = a_j]$ , and the transition probability from state  $i$  to state  $j$  is

$$q_{ij} = P[b(r, s+1) = a_j | b(r, s) = a_i]$$

which is nonzero if and only if  $(a_i, a_j) \in \Omega$ . The transition matrix  $T$  has entries  $q_{ij}$ . We describe a simple construction of a two-dimensional field such that any row (from left to right) and any column (from top to bottom) has the probability distribution of the chain  $M$ . We give the probability distribution for a finite  $N$  by  $L$  rectangular segment with indices  $1 \leq r \leq N$ ,  $1 \leq s \leq L$ . The distribution is defined by a causal

model in the sense that we give the conditional probability distribution of symbol  $B(r', s')$  conditioned on all symbols below,  $r > r'$ , and to the left,  $r = r'$  and  $s < s'$ .

Initially we let the left column,  $s = 1$ , and the bottom row,  $r = N$ , be specified by the chain  $M$ , i.e., the last symbol of the column,  $B(N, 1)$ , is used as the initial state for the row. We then specify the conditional probability of the random variables in row  $r$ ,  $B(r, s + 1)$ , given two of the neighbors,  $B(r, s)$  and  $B(r + 1, s + 1)$ . We prove that if the row  $r + 1$  is the Markov chain  $M$ , then the new row also has the distribution of  $M$ . Repeating this process the probability distribution of the entire rectangle is specified. It follows from the symmetry of the construction that the columns have the same probability distribution.

In order to describe the distribution of two successive rows we introduce a new Markov chain with the pair  $[B(r, s), B(r + 1, s)]$  as the state. In this chain we require that the transition probabilities satisfy

$$P[B(r + 1, s + 1) | B(r, s), B(r + 1, s)] \\ = P[B(r + 1, s + 1) | B(r + 1, s)]. \quad (2)$$

Thus, row  $r + 1$  can be extended independently of the past of row  $r$ , and we choose the conditional probabilities to coincide with the transition probabilities of  $M$

$$P[b(r + 1, s + 1) = a_u | b(r + 1, s) = a_v] = q_{vu}.$$

The definition of a symmetric Pickard field [3] similarly requires  $B(r, s + 1)$  to be independent of  $B(r + 1, s)$ , but in many cases these conditions cannot be satisfied. In the more general definition [4], [5], (2) is combined with a stationarity condition to give an independence condition in the opposite direction:  $B(r, s - 1)$  is independent of  $B(r + 1, s)$ . We proceed in a more direct manner letting the conditional probability of  $B(r, s + 1)$  be defined by

$$P[b(r, s + 1) = a_u | b(r, s) = a_v, b(r + 1, s) = a_x \\ b(r + 1, s + 1) = a_w] = \frac{q_{vu} q_{uw}}{\sum_t q_{vt} q_{tw}}. \quad (3)$$

This probability may be interpreted as the probability of a symbol in the Markov chain  $M$  given the previous and the following symbol. Thus, it does not depend directly on  $B(r + 1, s)$  but on  $B(r + 1, s + 1)$ . It follows from Lemma 1 that there is at least one nonzero term in the sum in the denominator of the right side of (3).

*Example 1 (Continued):* For the given set,  $\Omega$ , we can choose the transition matrix as

$$T = \begin{bmatrix} 1/4 & 1/2 & 0 & 1/4 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}.$$

The stationary probabilities for the Markov chain are  $(1/4, 1/8, 1/4, 3/8)$ . With these transition probabilities we can calculate the probability of the initial row and column from the first part of the example to be  $1/512$ . In row 3, we found that the symbols following  $D$  had to be  $EC$ , and thus, the conditional probabilities are 1. For the last symbol we get from (3)

$$P[b(3, 4) = O | b(3, 3) = C, b(4, 4) = E] \\ = \frac{1/2 \cdot 1/2}{1/2 \cdot 1/2 + 1/4 \cdot 1/4} = \frac{4}{5}.$$

In row 2 the only possibility is  $ODEC$ , and in row 1, the conditional probabilities of  $b(1, 2) = E$  and  $b(1, 3) = D$  are found from (2) as  $1/3$  and  $4/5$ . The probability of the outcome of the array is the product of these four factors,  $1/2400$ .

*Lemma 3:* The stationary distribution of the two-row chain is

$$P[b(r, s + 1) = a_u, b(r + 1, s + 1) = a_w] = Q_u q_{uw}.$$

*Proof:* Column 1 was chosen to have this distribution. Assuming it is true for column  $s$ , we get from (2) and (3)

$$P[b(r, s + 1) = a_u, b(r + 1, s + 1) = a_w] \\ = \sum_{v, x} Q_v q_{vx} q_{xw} \frac{q_{vu} q_{uw}}{\sum_t q_{vt} q_{tw}} = Q_u q_{uw}.$$

*Lemma 4:* The transition probabilities of row  $r$  satisfy

$$P[b(r, s + 1) = a_u | b(r, s) = a_v] = q_{vu}.$$

*Proof:* Similarly,

$$P[b(r, s + 1) = a_u | b(r, s) = a_v] \\ = \sum_{x, w} q_{vx} q_{xw} \frac{q_{vu} q_{uw}}{\sum_t q_{vt} q_{tw}} = q_{vu}.$$

We can now prove the following.

*Theorem 1:* If row  $r + 1$  is a Markov chain and the symbols in row  $r$  have conditional probabilities given by (3), row  $r$  is the same Markov chain.

*Proof:* In the two-row chain it follows from (2) that the probability distribution of row  $r + 1$  is

$$P[B(r + 1, 1)]P[B(r + 1, 2) | B(r + 1, 1)] \dots \\ P[B(r + 1, N) | B(r + 1, N - 1)]$$

in agreement with  $M$ . It follows from (3) and Lemmas 3 and 4 that the probability distribution of the triple  $(B(r, s), B(r, s + 1), B(r + 1, s + 1))$  is that of three consecutive variables in  $M$ , and thus

$$P[B(r, s) | B(r, s + 1), B(r + 1, s + 1)] = P[B(r, s) | B(r, s + 1)].$$

Thus, row  $r$  can be extended to the left independently of row  $r + 1$  to the right, and the transition probabilities are those of the Markov chain  $M$  taken in the reverse order. Consequently row  $r$  has the same distribution as  $M$ . Thus, when row  $r + 1$  and  $b(r, 1)$  are given, we can use (3) to specify the distribution of row  $r$  and get another copy of  $M$ .

The Markov property of the rows is a special case of the sufficient condition presented by Shtarkov and the author in [6]: The two-row Markov source provides a description of row  $r$  as a hidden state Markov source with  $B(r + 1, s + 1)$  as a hidden state variable. However, this variable in fact depends on only the observed state  $B(r, s + 1)$ , not on earlier symbols in row  $r$ .

Since the field is defined by the conditional probability given in (3), and this value depends only on the symbols immediately below and to the left, the causal model can also be interpreted as defining column  $s + 1$  conditioned on column  $s$ . Thus, the proof of Theorem 1 also shows that column  $s + 1$  is the Markov chain  $M$ .

A distribution on a finite array is said to be (locally) stationary [5] if the probability of any subset of symbols is invariant to any translation that keeps it entirely within the rectangle. Since the marginal distribution of any subset can be calculated from  $M$  and repeated applications of (3), we have

*Corollary 3:* The probability distribution is stationary.

It follows that the probability distribution defined here for a rectangle of arbitrary size can be extended to a stationary measure of a Markov random field in the entire first quadrant. Furthermore the field can be extended to an arbitrarily large part of the rest of the plane by translating the origin to a point in the third quadrant.

## IV. THE ENTROPY OF THE FIELD

Let the transition probabilities of the chain be  $q_{ij}$ . The two-step transition probabilities are found from  $T^2$  as

$$q'_{ik} = \sum_j q_{ij} q_{jk}.$$

*Theorem 2:* The entropy of the field is

$$R = \sum_i Q_i \sum_k q'_{ik} \sum_j -\frac{q_{ij} q_{jk}}{q'_{ik}} \log \left( \frac{q_{ij} q_{jk}}{q'_{ik}} \right). \quad (4)$$

*Proof:* We consider a rectangular segment of the field and apply the chain rule for the entropy. First the entropy of the left column and the bottom row is found from the entropy of  $M$ , denoted  $R_1$ , and the entropy of a single symbol  $R_0$ . Then the remaining symbols are taken row by row. Let  $U(r', s')$  be the set of variables with indices  $r > r'$  and  $r = r', s < s'$ . Writing the entropy as the expected value of the logarithm, it can then be expressed as

$$R_0 + (N + L - 2)R_1 - E \left\{ \sum_{r=1}^{N-1} \sum_{s=2}^L \log P[B(r, s) | U(r, s)] \right\}.$$

For all terms in the sum, it follows from (3) and Lemma 4 that

$$\begin{aligned} E \{ \log P[B(r, s) | U(r, s)] \} \\ &= E \{ \log P[B(r, s) | B(r, s-1), B(r+1, s)] \} \\ &= \sum_i Q_i \sum_j q_{ij} \sum_k q_{jk} \frac{q_{ij} q_{jk}}{q'_{ik}} \log \frac{q_{ij} q_{jk}}{q'_{ik}} \\ &= \sum_i Q_i \sum_k q'_{ik} \sum_j \frac{q_{ij} q_{jk}}{q'_{ik}} \log \frac{q_{ij} q_{jk}}{q'_{ik}}. \end{aligned}$$

The conditional entropy of  $B(r, s+1)$  is found from (3). The expected value thus, equals  $R$  as given by (4), and the entropy of the field becomes

$$\lim_{N, L \rightarrow \infty} \left[ \frac{R_0 + (N + L - 2)R_1 + (N - 1)(L - 1)R}{NL} \right] = R.$$

*Example 1 (Continued):* The entropy is found from (4). The only triples that give positive contributions to the entropy are  $CCC$ ,  $CEC$ ,  $COD$ ,  $CED$ ,  $CCE$ ,  $COE$ ,  $ECE$ , and  $EDE$ . We get

$$\begin{aligned} R &= \frac{3}{64} H\left(\frac{1}{3}\right) + \frac{3}{32} H\left(\frac{1}{3}\right) + \frac{5}{64} H\left(\frac{1}{5}\right) + \frac{15}{64} H\left(\frac{1}{5}\right) \\ &= \frac{9}{64} H\left(\frac{1}{3}\right) + \frac{5}{16} H\left(\frac{1}{5}\right) = 0.352. \end{aligned}$$

*Lemma 5:* For a constrained array,  $R > 0$  if and only if the combinatorial entropy is positive.

*Proof:* The first part follows from Lemma 2 and (3) since all permissible outcomes are assigned a positive probability. The second part is obvious.

*Lemma 6:* The entropy of  $M$  is an upper bound on the entropy of any 2-D field with rows and columns satisfying this distribution. If any field reaches this bound,  $R$  has the same value.

*Proof:* Since the distribution of pairs  $B(r, s)$ ,  $B(r, s+1)$  are given by  $M$ , the entropy of the field cannot exceed the conditional entropy of  $B(r+1, s)$  given  $B(r, s)$ , which is the entropy of  $M$ . However, for the field to reach this entropy,  $B(r, s+1)$  and  $B(r+1, s+1)$  have to be independent given  $B(r, s)$ . If we let the distribution of this triple define a second order Markov chain, it must again have the same entropy, since the entropy is at least that of the field and cannot exceed

the entropy of  $M$ . The only possible structure is a chain with period 2, and we provide an example.

*Example 2:* The concepts presented in this section can be illustrated by considering the alphabet  $A = \{C, D, E\}$  and the following constraints:

- 2.1:  $\Omega = \{(CD), (DC), (DE), (EC)\}$ . The Markov chain can have an entropy of 0.4, but only trivial arrays exist, and thus, the entropy of the 2-D field is 0.
- 2.2:  $\Omega = \{(CD), (DC), (CE), (EC)\}$ . Both the Markov chain and the field have maximal entropy 0.5, since the chain alternates between the symbol  $C$  and one of the other symbols. We get the same result from Theorem 2.
- 2.3:  $\Omega = \{(CD), (DC), (CC)\}$ . This is the well-known binary “hard-square” field. The largest value of the entropy for the 2-D field is known with high precision as 0.5879 [7], but the distribution has none of the simple properties described here. With the probability distribution described in this correspondence and  $P[D] = 0.222$  we get  $R = 0.5831$ .

Lemma 5 and 6 and Example 2 indicate that the simple fields constructed here have a relatively high entropy, but that they reach the maximal values only in extreme cases. However, considering the other approximations involved in data compression or encoding, these fields may provide a useful starting point.

*Example 3:* Consider an alphabet of at least five symbols,  $a_j$ ,  $0 \leq j \leq J$ , where a transition can occur only from state  $j$  to state  $j$  or  $j \pm 1$  modulo  $J+1$ . (One may think of the 2-D field as an image where the color changes only incrementally in each direction). In one dimension the entropy is maximized by taking a symmetric probability assignment. If the probability of staying in the same state is  $1 - 2\alpha$  and the probability of changing in either direction is  $\alpha$ , the entropy for a one dimensional chain is clearly  $R_1 = H(2\alpha) + 2\alpha$ . Assuming all symbols equally likely, we find for the field  $q'_{ik} = (1 - 2\alpha)^2 + 2\alpha^2$  for  $k = i$  and  $2\alpha(1 - 2\alpha)$  for  $k = i \pm 1$ . From (4) we find the entropy

$$R = (1 - 4\alpha + 6\alpha^2) \left( H\left(\frac{2\alpha^2}{1 - 4\alpha + 6\alpha^2}\right) + \frac{2\alpha^2}{1 - 4\alpha + 6\alpha^2} \right) + 4\alpha(1 - 2\alpha).$$

The largest value is exactly 1 bit for  $\alpha = 1 - \sqrt{2}/2$ . If we eliminate the transitions between the extreme states (the field may model a surface whose amplitude changes only in steps of at most 1), the entropy in one dimension is reduced, but in two dimensions we can get a slightly larger value. Thus, if only the allowed transitions are given, the maximal entropy occurs for a nonunique asymmetric probability assignment. A simple chain with near maximal entropy uses only three states with

$$T = \begin{bmatrix} 1/3 & 2/3 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 2/3 & 1/3 \end{bmatrix}.$$

From (4) we find the entropy as

$$R = \frac{1}{3} H\left(\frac{3}{7}\right) + \frac{4}{21} + \frac{25}{42} H\left(\frac{2}{5}\right) \approx 1.097.$$

## V. MODIFICATIONS

In this section, we briefly describe how some more general fields, which have been discussed in the literature, can be understood in terms of the methods presented here.

If the independence condition (2) can be satisfied, the constraint on a two by two square may still require that the transition probabilities (3)

be modified to depend directly on  $B(r+1, s)$ . The following example, which was also presented in [7] illustrates such a case.

**Example 4:** Let the rows be a Markov chain with three symbols and transition matrix

$$T = \begin{bmatrix} 1-2p & p & p \\ p & 1-2p & p \\ p & p & 1-2p \end{bmatrix}.$$

We obtain a tiling of the plane with rectangles of three colors such that no adjacent rectangles have the same color by imposing the constraint that if three of the variables in a two by two square have the same color, the fourth must have the same color. Thus, the conditional probabilities (3) must be modified to satisfy this constraint, but we want to leave the marginal distribution  $P[B(r, s), B(r, s+1), B(r+1, s+1)]$  unchanged. For

$$P[b(r, s) = b(r+1, s) = b(r+1, s+1) = C]$$

$b(r, s+1)$  has to be  $C$ , and the two triples have the same probability. There is also no problem in the case of

$$P[b(r, s) = C, b(r, s+1) = D, b(r+1, s+1) = C]$$

since the distribution is maintained by any choice of the transition probability,  $(0, 1-q, q)$  as long as the symmetry in the symbols is preserved. However, the conditional probabilities

$$\begin{aligned} P[b(r, s+1) = X | b(r, s) = C, \\ b(r+1, s) = C, b(r+1, s+1) = D] \\ P[b(r, s+1) = X | b(r, s) = C, \\ b(r+1, s) = D, b(r+1, s+1) = E] \end{aligned}$$

can be written as  $(0, 1-q_1, q_1)$ , and  $((1-q_2)/2, q_2, (1-q_2)/2)$ , where the parameters must satisfy

$$2q_1(1-2p) = p(1-q_2).$$

Now the distribution of the sizes of the rectangles can be controlled by a suitable choice of the parameters, and the entropy can be explicitly calculated as

$$R = 2p^2 H(q) + 2p^2 (H(q_2) + 1 - q_2) + 4p(1-2p)H(q_1).$$

If the constraint on diagonal elements prevents (2) from being satisfied, or if a distribution is desired where transition probabilities depend on the other symbol in the diagonal, the state variables in the two-row chain can be shifted to  $B(r+1, s+1)$  and  $B(r, s)$ . Each row is still assumed to be the original chain  $M$ , but the state variables are two consecutive symbols in a different Markov chain,  $M'$ , which describes the diagonal with increasing values of  $r$  and  $s$ . We could recover the construction of the previous section by taking  $M'$  to have transition matrix  $T = T^2$ , but it is also possible to choose a different process. As discussed in [6], the condition for combining the two chains is that the transition probabilities commute (in particular they must have the same stationary distribution).

**Example 5:** In [8] this idea was used to increase the entropy of the hard-square model compared to the value given in Example 2. With the same notation as in the Example 2,  $P[D] = p$  and the additional parameter  $P[b(r+1, s+1) = b(r, s) = D] = p_2$ , let the transition matrix of the diagonal process be

$$T = \begin{bmatrix} \frac{1-2p+p_2}{1-p} & \frac{p-p_2}{p} \\ \frac{p-p_2}{1-p} & \frac{p_2}{p} \end{bmatrix}$$

and we get the entropy

$$R = \frac{(1-2p+p_2)(1-2p)}{1-p} H\left(\frac{(p-p_2)(1-p)}{(1-2p+p_2)(1-2p)}\right) + \frac{(1-2p+p_2)p}{1-p} H\left(\frac{p_2(1-p)}{(1-2p+p_2)p}\right)$$

which has maximal value  $R = 0.5873$ .

In information theory, Markov sources or functions of Markov chains are often used to describe more complex constraints, such as limits on runlengths, in sequences with small alphabets. If such a constraint is imposed on the rows and columns of a 2-D field, the result is usually a different set of transitions in the two directions. Thus, the construction does not apply to such fields. On the other hand, the increments of the field considered in Example 3 would provide a ternary field with bounded sum variation as considered in [9].

## VI. CONCLUSION

The construction presented here provides a natural link to the use of Markov chains in information theory. It also suggests a simple encoding of information. The conditional probabilities can be chosen to approximate the statistics of observed data, or they can be chosen to achieve large, although usually not maximal, entropy for a set of constraints.

## ACKNOWLEDGMENT

The author would like to acknowledge many suggestions from the reviewers, which helped clarify the presentation.

## REFERENCES

- [1] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for digital recorders," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2260–99, Oct. 1998.
- [2] C. E. Shannon, "A mathematical theory of communications," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.
- [3] D. K. Pickard, "A curious binary lattice process," *J. Appl. Prob.*, vol. 14, pp. 717–731, 1977.
- [4] —, "Unilateral Markov fields," *Adv. Appl. Probab.*, vol. 12, pp. 655–671, 1980.
- [5] F. Champagnat, J. Idier, and Y. Goussard, "Stationary Markov random fields on a finite rectangular lattice," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2901–2916, Nov. 1998.
- [6] J. Justesen and Y. M. Shtarkov, "Simple models of two-dimensional information sources and codes," in *Proc. IEEE Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 412.
- [7] S. Forchhammer and J. Justesen, "Entropy bounds on constrained two-dimensional fields," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 118–127, Jan. 1999.
- [8] R. M. Roth, P. H. Siegel, and J. K. Wolf, "Efficient coding schemes for the hard-square model," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1166–1176, Mar. 2001.
- [9] J. Justesen, "Bounds on constrained codes in one and two dimensions," in *Codes, Curves, and Signals: Common Threads in Communications*, A. Vardy, Ed. Boston, MA: Kluwer Academic, 1998.